

人工智能中的联结主义和符号主义

顾险峰

纽约州立大学石溪分校计算机系, 美国纽约 11794

摘要 2016年3月,谷歌AlphaGo击败了围棋九段李世石,举世震惊。时人有为人工智能的发展欢呼雀跃者,有为人类前途命运忧心忡忡者,有对机器蛮力不屑一顾者,有对人类失去优越感而沮丧彷徨者。本文探讨人工智能发展的主要方向,分析人工智能在目前的局限和未来的潜力,通过对人类脑神经认知和人工神经网络认知进行对比,从而对人工智能有一个公正客观而又与时俱进的认识。

关键词 人工智能;联结主义;符号主义

人类的智能主要包括归纳总结和逻辑演绎,对应着人工智能中的联结主义(如人工神经网络,图1为由人工神

神经网络生成的 Inceptionist 图像)和符号主义(如吴文俊方法)。人类大量的视觉听觉信号的感知处理都是下意识的,

基于大脑皮层神经网络的学习方法;大量的数学推导、定理证明是有强烈主观意识的,是基于公理系统的符号演算方法。



图1 由人工神经网络生成的 Inceptionist 图像



图2 三维曲面到平面的保角映射

1 联结主义

David Hunter Hubel 和 Torsen Wiesel 共同获得了 1981 年诺贝尔生理学或医学奖。1959 年, Hubel 和 Wiesel 在麻醉的猫的视觉中枢上插入微电极后,在猫的眼前投影各种简单模式,观察猫的视觉神经元的反应。他们发现,猫的视觉中枢中有些神经元对于某种方向的直线敏感,另外一些神经元对于另外一种方向的直线敏感;某些初等的神经元对于简单模式敏感,另外一些高级的神经元对于复杂模式敏感,并且其敏感度和复杂模式的位置与定向无关。证明了视觉中枢系统具有由简单模式构成复杂模式的功能,也启发了计算机科学家发明人工神经网络。

1.1 层次特征(hierarchical features)

后来,通过对猴子的视觉中枢的解剖,将猴子的大脑皮层曲面平展在手术台表面上,人们发现从视网膜到第一级视觉中枢的大脑皮层曲面的映射(retinotopic mapping)是保角映射(conformal mapping)^[1]。如图2所示,保角变换的最大特点是局部保持形状,但

收稿日期:2016-03-14

作者简介:顾险峰,教授,研究方向为计算共形几何及其在计算机视觉、图形学、网络、医学图像上的应用,电子信箱:gu@cs.stonybrook.edu

引用格式:顾险峰. 人工智能中的联结主义和符号主义[J]. 科技导报, 2016, 34(7): 20-25; doi: 10.3981/j.issn.1000-7857.2016.07.002

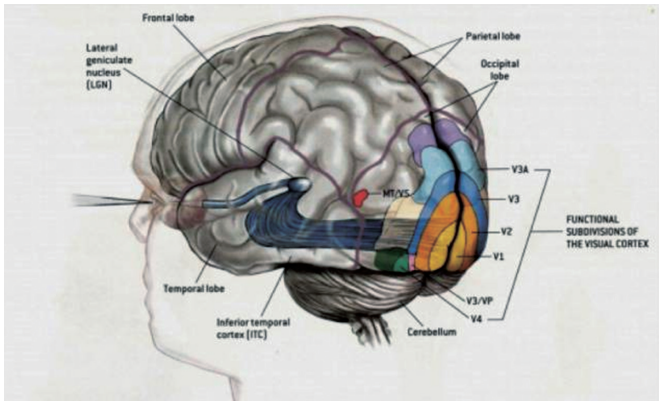


图3 大脑皮层的视觉中枢,视觉信号的传导途径: 视网膜, LGN, V1, V2, V3, V4, V5等

是忽略面积大小,这说明视觉处理对于局部形状非常敏感。

人们逐步发现,人类具有多个视觉中枢,并且这些视觉中枢是阶梯级联,具有层次结构。人类的视觉计算是一个非常复杂的过程。如图3所示,在大脑皮层上有多个视觉功能区域(V1至V5等),低级区域的输出成为高级区域的输入。低级区域识别图像中像素级别的局部的特征,例如边缘折角结构,高级区域将低级特征组合成全局特征,形成复杂的模式,模式的抽象程度逐渐提高,直至语义级别。

忽略色彩、纹理、光照等局部细节,侧重整体模式匹配和上下文关系,并可以主动补充大量缺失信息。

近年来,深度学习技术的发展,使得人们能够模拟视觉中枢的层级结构,考察每一级神经网络形成的概念。图5显示了一个用于人脸识别的人工神经网络经过训练后得到的各层特征。底层网络总结出各种边缘结构,中层网络归纳出眼睛、鼻子、嘴巴等局部特征,高层网络将局部特征组合,得到各种人脸特征。这样,人工神经网络佐证了视觉中枢的层次特征结构。



图4 毕加索的Guernica

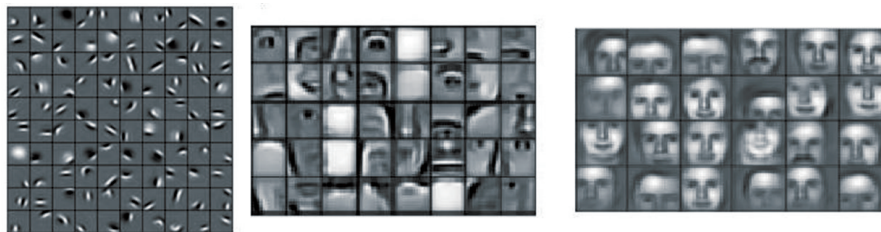


图5 深度学习神经网络学习得到的不同层次的特征(by Andrew Ng)

如图4所示,毕加索的名画格尔尼卡(Guernica)中充满了抽象的牛头马面,痛苦嚎哭的人脸,扭曲破碎的肢体。人们从图4中可以毫不费力地辨认出这些夸张的几何形体。其实,图4中大量信息丢失,但是提供了足够的整体模式。由此可见,视觉高级中枢

1.2 专用和通用(specific vs. general)

人工神经网络在20世纪80年代末、90年代初达到巅峰,随后迅速衰落,其中一个重要原因是深度神经网络的发展严重受挫。人们发现,如果网络的层数加深,那么最终网络的输出结果对于初始几层的参数影响微乎其微,整个网络的训练过程无法保证收敛。

同时,人们发现大脑具有不同的功能区域(图6),每个区域专门负责同一类的任务,例如视觉图像识别、语音信号处理和文字处理等。并且在不同的个体上,这些功能中枢在大脑皮层上的位置大致相同。在这一阶段,计算机科学家为不同的任务发展出不同的算法。例如,为了语音识别,人们发展了隐马尔科夫链模型;为了人脸识别,发展了Gabor滤波器、SIFT特征提取算子、马尔科夫随机场的图模型。因此,在这个阶段,人们倾向于发展专用算法。

而脑神经科学的几个突破性进展使人们彻底改变了看法。2000年左右, Jitendra Sharma在《Nature》上发表文章,报告了一个令人耳目一新的实验^[2]。Sharma把幼年鼯鼠的视觉神经和听觉神经剪断,交换后接合,眼睛接到了听觉中枢,耳朵接到了视觉中枢。鼯鼠长大后,依然发展出了视觉和听觉。这意味着大脑中视觉和听觉的计算方法是通用的。2009年, Vuilleme和 Cuisinier为盲人发明了一套装置^[3],将摄像机的输出表示成二维微电极矩阵,放在舌头表面。盲人经过一段时间的学习训练,可以用舌头“看到”障碍物。2011年,人们发现许多盲人独自发展出一套“声纳”技术,他们可以通过回声来探测并规避大的障碍物。Thaler等研究表明,他们“声纳”技术采用的并不是听觉中枢,而是原来被废弃的视觉中枢。

种种研究表明,大脑实际上是一台“万用学习机器”(universal learning machine),同样的学习机制可以用于完全不同的应用。人类的DNA并不提供各种用途的算法,而只提供基本的普适学习机制,人的思维功能主要是依赖于

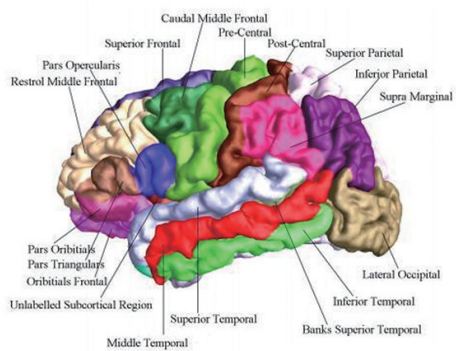


图6 大脑皮层的不同功能区域

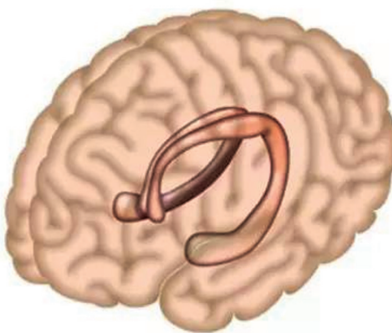
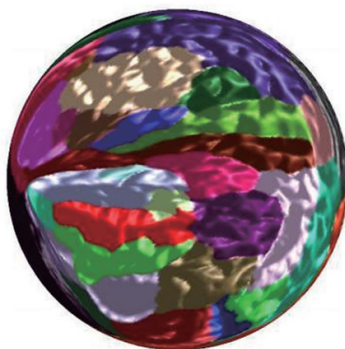


图7 大脑的海马体 (Hippocampus)

学习所得。后天的文化和环境决定了一个人的思想和能力。换句话说,学习的机制人人相同,但是学习的内容决定了人的思想。

1.3 可塑性(plasticity)

人的大脑具有极强的可塑性,许多功能取决于后天训练。例如,不同民族语言具有不同的元音和辅音,阿拉伯语最为复杂,日语相对简单。出生不久的婴儿可以辨别听出人类能够发出的所有元音和辅音,但是在5岁左右,日本幼儿已经听不出很多阿拉伯语中的音素了。同样,欧洲人可以容易地辨认本民族面孔,但是非常容易混淆亚洲人面孔。人们发现,如果大脑某个半球的一个区域受损,产生功能障碍,随时间流逝,另一半球的对称区域会“接替”受损区域,掌管相应功能。这些均表明大脑神经网络具有强烈的可塑性。

长期以来,人们倾向于认为大脑神经网络通过学习和训练,其联结复杂度逐渐增加,愈来愈多的联结建立起来。近期,一些神经科学家提出了相反的看法。他们观察到,婴儿睡觉时,如果有剪刀掉到地上,婴儿的应激反应是全身的,而相对成熟的儿童的应激反应只集中在局部肌肉上面。他们找到一些证据表明,婴儿的某些神经网络是全联通图,随年龄增长和学习训练的积累,许多神经联结会自行断开,从而形成简化的网络。

1.4 梦的解析(dream)

大脑学习算法的普适性和可塑性一直激励着计算机科学家不懈地努力

探索。历史性的突破发生在2006年左右。Geoffrey Hinton、Yann LeCun 和 Yoshua Bengio 3位计算机科学家突破了深度学习的技术瓶颈,引领了深度学习的浪潮。相比以前的状况,主要的技术突破在于以下几点:优化方法的改进,更加简单的优化方法,特别是随机梯度下降方法的应用;使用非监督数据训练模型以达到特征自动提取;使用越来越大的数据集;深度神经网络和大数据训练需要巨大的计算能力,GPU的普遍使用解决了这一迫切要求等。现在,深度学习方法突飞猛进,在图像识别(image registration)、文本处理(text)、语音处理(speech)等领域的基本问题上,都已经超过了传统方法。在图像识别领域,2015年深度学习方法的识别准确率已经达到人类的水平^[4]。

同时,对于深度学习神经网络的理解加深了人类对于自身智力活动的理解。长期以来,人们对于梦境一直没有很好的理解。一种解释是,如图7所示,大脑中有一对海马体(Hippocampus),它们和人类的长期记忆有关。如果把大脑比喻成一个数据库,那么海马体就像是索引。如果海马体有问题,那么许多存入的记忆无法被取出,同时也无法形成新的记忆。每天晚上,海马体将当天形成的短暂记忆加工成长期记忆,在这一过程中,

就形成了梦。海马体和其他神经中枢相连,处理其他中枢已经处理好的数据,形成新的编码。海马体和视觉与听觉中枢直接相连,因此在梦中能够看到并且听到;但海马体和嗅觉中枢不相连,因此在梦中无法闻到气味。在梦中经常能够看到平时看不到的奇诡景象,可用深度学习方法加以模拟解释。

实际上,视觉处理的过程并不只是从低级向高级传递的单向过程,高级中枢可以向低级中枢发出反馈信息,最明显的例子是高级中枢可以决定低级中枢的“注意力”和“焦点”。当看到模糊不清或一时无法辨别的图像时,高级中枢会产生各种概率上合理的解释,并且由这种猜测先入为主地影响低层中枢的判断,从而产生错觉。如图8^[5]所示,可以用深度学习的神经网络模拟这种先入为主的现象。输入一幅白噪声,本身没有任何有意义的信息。网络由于随机涨落,在某一瞬间倾向认为图中有香蕉。由此,优化这幅图像,使得识别香蕉的高层神经元兴奋,如此得到的图像果真看起来像香蕉。

图9^[5]显示了另外一个例子,输入一幅羚羊图像,神经网络的低级反馈加

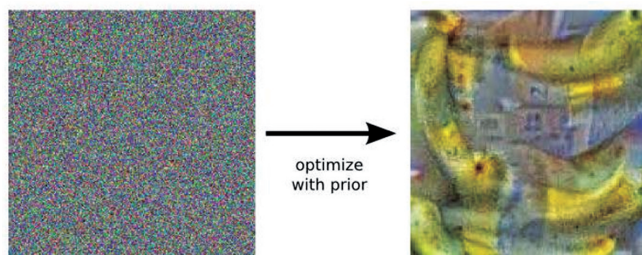


图8 先入为主产生的错觉

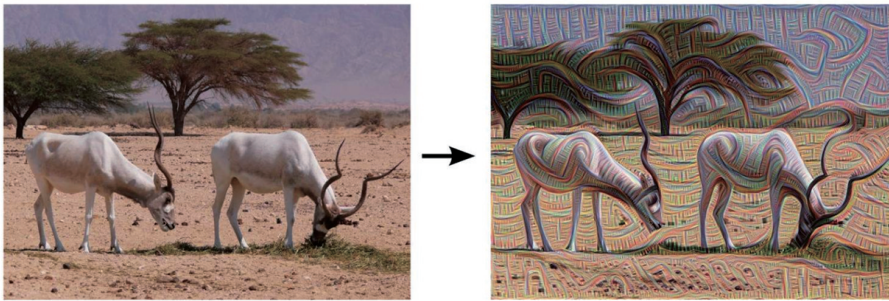


图9 图像加入低级特征反馈

到图像上,看到许多边缘和定向的模式出现在场景里。

许多孩子喜欢仰望蓝天白云,并且用自己丰富的想象力看到了各种奇妙的幻象。如图 10^[9]所示,将一幅蓝天白云的图像作为输入,用一个识别动物的深度学习神经网络加以处理,将高层神

经元的认知模式作为反馈,优化原始图像,结果可以看到各种山海经中才会出现的神兽:身着铠甲的将军狗、猪蜗牛、骆驼鸟、狗鱼。人在做梦时,高层神经元对于低层神经元发出各种反馈,低层神经元将图像依照高层的意图进行诠释幻化,视觉幻象由此产生。

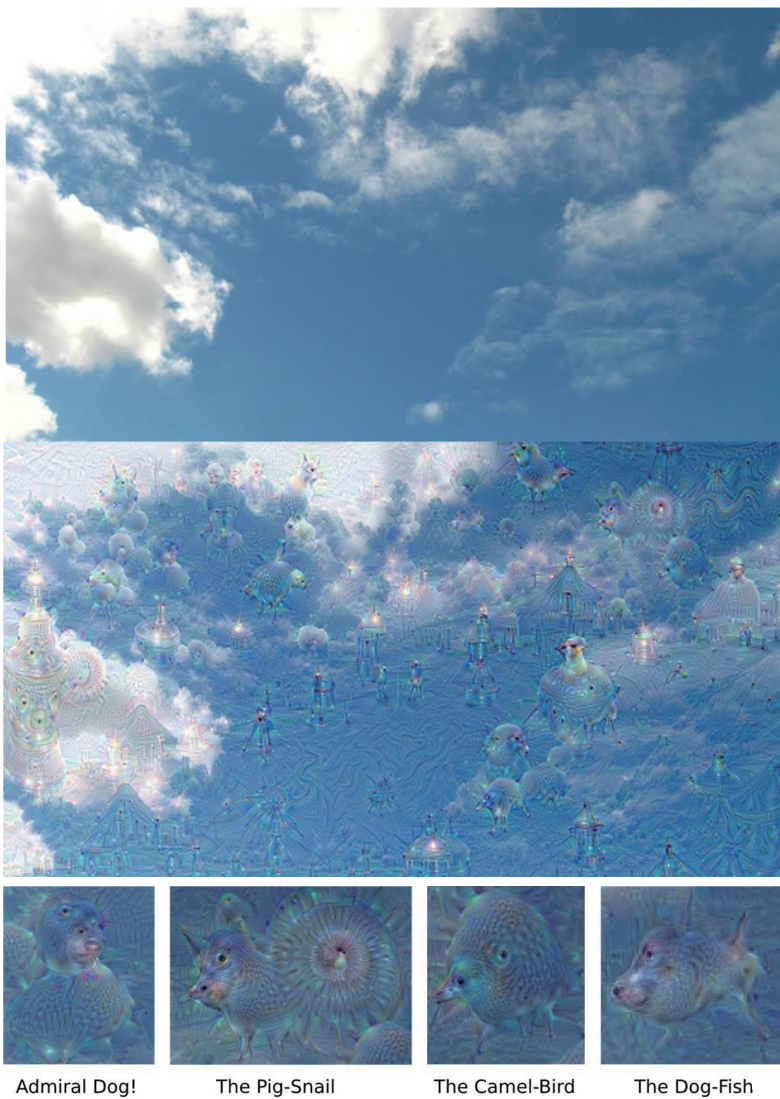


图 10 白云苍狗的机器学习解释

1.5 美学(aesthetics)

很久以来,人们倾向于认为机器可以理解人类的逻辑思维,但却无法理解人类的丰富感情,更无法理解人类的美学价值,当然机器也就无法产生具有美学价值的作品。事实胜于雄辩,AlphaGo 对局李世石下出石破天惊的一步,棋圣聂卫平向 AlphaGo 的下法脱帽致敬,这说明深度学习算法已经能够自发创造美学价值。许多棋手在棋盘方寸间纵横一生,所追寻的就是美轮美奂的神机妙手。如此深邃优美,玄奥抽象,一夜间变成了枯燥平淡的神经元参数,这令许多人心生幻灭。

其实,在视觉艺术领域,人工神经网络已经可以将一幅作品的内容和风格分开,同时向艺术大师学习艺术风格,并把艺术风格转移到另外的作品中,用不同艺术家的风格渲染同样的内容,如图 11 所示(详见:<https://github.com/jcjohnson/neural-style>)。

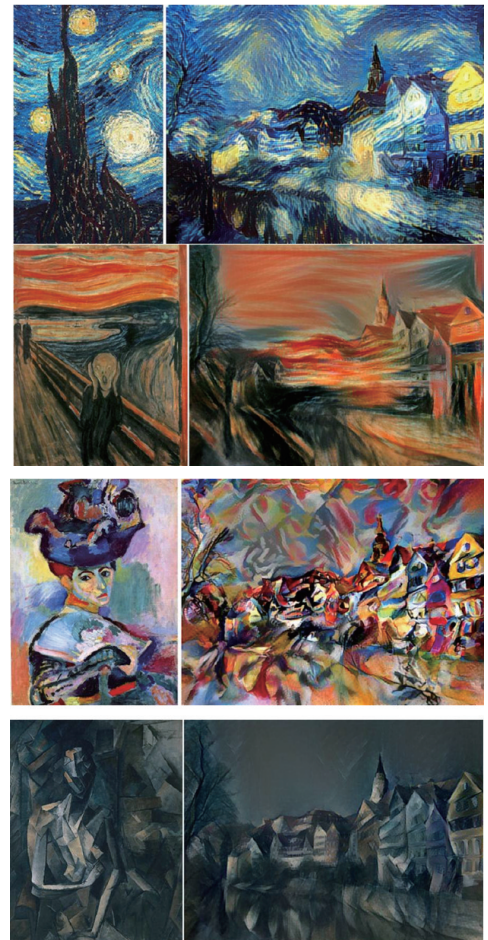


图 11 神经网络能够自动学习艺术风格,并用不同的风格渲染同样的内容

这意味着人工智能可以精确量化原本许多人文科学中模糊含糊的概念,例如特定领域中的“艺术风格”,博弈中的“棋风”,并且使这些只可意会、无法言传的技巧风格变得朴实无华,容易复制和推广。

2 符号主义

古希腊人将欧几里得几何归纳整理成欧几里得公理体系,整个宏伟的理论大厦奠基于几条不言自明的公理,完全由逻辑构造出来,美轮美奂,无懈可击。这为整个人类科学发展提供了一套标准的范式。后来,牛顿编撰鸿篇巨著《自然哲学的数学原理》也遵循公理体系的范式,由公理到定义、引理、定理,再到推论。人类的现代数学和物理知识最终都被系统化整理成公理体系,比如爱因斯坦的广义相对论也是遵循公理体系的范式。当然也存在例外,虽然量子理论已经为人类科技带来天翻地覆的革命,但是量子理论的公理体系目前还没有建立起来。符号主义的主要思想就是应用逻辑推理法则,从公理出发推演整个理论体系。

人工智能中,符号主义的一个代表就是机器定理证明,吴文俊先生创立的吴文俊方法是其巅峰之一。目前基于符号计算的机器定理证明的理论根基是希尔伯特定理:多元多项式环中的理想都是有限生成的。首先将一个几何命题的条件转换成代数多项式,同时把结论也转换成多项式,然后证明条件多项式生成的根理想包含结论对应的多项式,即将定理证明转换为根理想成员判定问题。一般而言,多项式理想的基底并不唯一,Groebner基方法和吴文俊方法可生成满足特定条件的理想基底,都可自动判定理想成员问题。因此理论上代数范畴的机器定理证明可以被完成,但实际中这种方法有重重困难。

首先,从哲学层面上讲,希尔伯特希望用公理化方法彻底严密化数学基础。哥德尔证明了对于任何一个包含算术系统的公理体系,都存在一个命题,其真伪无法在此公理体系中判定。换言之,这一命题的成立与否都与此公

理体系相容。这意味着我们无法建立包罗万象的公理体系,无论如何,总存在真理游离在有限公理体系之外;另一方面,这也意味着对于真理的探索过程永无止境。

其次,从计算角度而言,Groebner基方法和吴文俊方法所要解决的问题的本质复杂度都是超指数级别的,即便对于简单的几何命题,其机器证明过程都可能引发存储空间指数爆炸,这揭示了机器证明的本质难度。吴文俊方法的成功有赖于大多数几何定理所涉及的代数计算问题是有结构的,因而可以快速求解。

第三,能够用理想生成的框架证明的数学命题,其本身应该是已经被代数化了。例如所有的欧几里得几何命题、初等的解析几何命题。微分几何中的许多问题的代数化,本身就是非常具有挑战性。例如黎曼流形的陈省身-高斯-博内定理:流形的总曲率是拓扑不变量。如果没有嘉当发明的外微分和活动标架法,这一定理的证明无法被代数化。拓扑学中的许多命题的代数化本身也是非常困难的,比如众所周知的布劳威尔不动点定理:用咖啡勺缓慢均匀搅拌咖啡,然后抽离咖啡勺,待咖啡静止后,必有一个分子,其搅拌前和搅拌后的位置重合。这一命题的严格代数化是一个非常困难的问题。吴先生的高足,高小山研究员突破的微分结式理论,系统地将这种机器证明方法从代数范畴推广到微分范畴^[6]。

最后,机器定理证明过程中推导出的大量符号公式,人类无法理解其内在的几何含义,无法建立几何直觉。而几何直觉和审美,实际上是指导数学家在几何天地中开疆拓土的最主要的原则。机器无法抽象出几何直觉,也无法建立审美观念,因此虽然机器定理证明经常对于已知的定理给出令人匪夷所思的新颖证明方法,但是迄今为止,机器并没有自行发现深刻的未知数学定理。

比如,人类借助计算机完成了地图四色定理的证明,但是对于这一证明的意义一直富有争议。首先,这种暴力证

明方法没有提出新的概念、新的方法;其次,这个证明没有将这个问题和其他数学分支发生深刻内在的联系。数学中,命题猜测的证明本身并不重要,真正重要的是证明所引发的概念思想,内在联系和理论体系。因此,许多人认为地图四色定理的证明实际上“验证”了一个事实,而非“证明”了一个定理。目前,机器定理证明的主流逐渐演变成机器验证。

因此,和人类智慧相比,人工智能的符号主义方法依然处于相对幼稚的阶段。

3 展望

虽然人工智能取得了突破性进展,但是它还处于婴幼儿时期。联结主义的方法虽然摧枯拉朽、无坚不摧,但是依然没有坚实的理论基础。通过仿生学和经验积累得到的突破,依然无法透彻理解和预测。简单的神经网络学习机制加上机器蛮力,能否真正从量变到质变,这需要时间检验。

围棋是信息完全博弈游戏的巅峰,但不是人类智力的巅峰。理性思维的巅峰还是数学物理理论的创立。许多抽象的数学定理,本身的描述已经概念嵌套概念,并且在现实物理世界中找不到示例,因此机器学习方法不再适用。比如下面的命题:高斯曲率处处为负的封闭曲面无法嵌入在三维欧式空间中。这一命题的证明无法用目前符号计算的方法,也无法用机器学习的方法。如果有朝一日,数学家开始对人工智能脱帽致敬,那么人类应该开始警醒。

人类的孩子需要花费十数年来学习,在这漫长的学习过程中,师生和同学中的社会交往起到了至关重要的作用。目前人工智能的程序还是各自训练,彼此之间没有交互。如果人工智能程序间能够建立社会交往,彼此交流学习内容和心得,那么人工智能将会产生新的飞跃。比如一群谷歌的Atlas机器人(图12),装备了AlphaGo的大脑,彼此自发地进行篮球游戏。通过拷贝神经网络的参数,它们可以彼此迅速交换



图12 谷歌旗下的机器人

学习成果。通过对抗演练,它们能够迅速提高手眼协调,团队协作,则很快有望进入NBA。

也许,在不久的将来,人类和人工智能机器人和睦共处,共同学习提高。在纽约曼哈顿中央公园,有机器人在辛勤地劳作,也有人遛着阿尔法狗(机器人宠物)。街角,有个Atlas在和人类攀谈,切磋着各种幽默笑话。不远处的长椅上另一个Atlas默默地坐着,它头戴虚拟现实的眼镜,在孜孜不倦地学习着人类历史……

参考文献(References)

- [1] Brewer A A, Liu J V, Wade A R, et al. Visual field maps and stimulus selectivity in human ventral occipital cortex[J]. Nature Neuroscience, 2005; 8(10): 1411-1411.
- [2] Sharma J, Angelucci A, Sur M. Induction of visual orientation modules in auditory cortex[J]. Nature, 404(6780): 841-847.
- [3] Vuillerme N, Cuisinier R. Sensory supplementation through tongue electro tactile stimulation to preserve head stabilization in space in the absence of vision[J]. Investigative Ophthalmology & Visual Science, 2009, 50(1): 476-481.
- [4] He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition[R]. arXiv:1512.03385.
- [5] Inceptionism: Going Deeper into Neural Networks[EB/OL]. [2015-06-01]. <http://googleresearch.blogspot.com/2015/06/inceptionism-going-deeper-into-neural.html>.
- [6] Gao X S, Li W, Yuan C M. Intersection theory in differential algebraic geometry: Generic intersections and the differential chow form[J]. Transactions of the American Mathematical Society, 2013, 365(9): 4575-4632.

(责任编辑 刘志远)